

Title	孤立者検出のための立食形式パーティー映像のハンド アノテーション分析
Author(s)	山内, 賢幸; 坊農, 真弓; 相原, 健郎; 西本, 一志
Citation	情報処理学会研究報告: ヒューマンコンピュータイン タラクション, 2010-HCI-137(10): 1-8
Issue Date	2010-03-12
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/9498
Rights	<p>社団法人 情報処理学会, 山内賢幸, 坊農真弓, 相原健郎, 西本一志, 情報処理学会研究報告: ヒューマンコンピュータインタラクション, 2010-HCI-137(10), 2010, 1-8. ここに掲載した著作物の利用に関する注意: 本著作物の著作権は(社)情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 Notice for the use of this material: The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) Information Processing Society of Japan.</p>
Description	

孤立者検出のための立食形式パーティー映像のハンドアノテーション分析

山内賢幸[†] 坊農真弓^{††} 相原健郎^{††} 西本一志[†]

学会の懇親会や結婚式の披露宴などの立食形式パーティーで会話の輪から孤立している人がいる。この会話の輪に入ることが出来ない状態の人を「孤立者」と呼ぶ。本研究では立食形式パーティーの映像を使いハンドアノテーションを行う。そしてそこから得られたアノテーションデータを用いて孤立者の検出を行う。今回提案する検出手法は会話集団を見つけ、それ以外の人を孤立者とする方法である。人が映像を見て直感的に孤立者と判断したデータと提案手法で得られた結果との比較を行い、どれだけ孤立者が検出できたのかを明らかにする。またその結果から今後の課題についても検討を行う。

A hand annotation method for video analysis of a buffet party ~Toward detection of wallflowers~

Yoshiyuki Yamauchi[†] Mayumi Bono^{††}
Kenro Aihara^{††} and Kazushi Nishimoto[†]

In this paper, we discuss a hand annotation method for video analysis of a buffet party toward detection of wallflowers. There are often some wallflowers in the banquet or the wedding party. We try to find out the wallflowers thorough understanding various participation roles, ex. speaker, addressee, side participant, by the multimodal analysis of integrating several hand annotation data, i.e., speech, gaze direction body orientation and standing positions. We compare the wallflowers found out by our algorithm and that found out by people's intuitions to investigate the performance of our algorithm.

1. はじめに

古くから立食形式のパーティーが世界中で行われてきた。我々にとって身近なものとしては、結婚披露宴や学会の懇親会などがあげられる。このような場で、ぼつんと一人きりになってしまっている人を見かけたことはないだろうか。欧米などでは、ダンスパーティーで、誰にも相手にされずに壁際に佇んでいる若い女性のことを“壁の花”(wall flower)と呼んでいる。この言葉が示しているように、実際にパーティー会場において会話の輪に入ることの出来ない孤立している状態の人が存在している。本研究において、この会話の輪に入ることが出来ない状態の人を「孤立者」と呼ぶ。

それでは、いったいどのような状態の人が「孤立者」として定義でき、またセンシング技術などによってそれらの人々を検出できるのだろうか。従来、各種センサーを利用して位置情報を得たり、マイクから発話情報を得たりといったインタラクション理解の研究が数多くなされてきた。特に角ら[1][2]のように人に複数のセンサーをつけ、そこから記録された大量のデータからインタラクションのコーパスを構築する手法などがあげられる。しかしながら、ハードウェアを利用したセンサー主体のインタラクション理解の発想では、実際の人間の会話や行動、ジェスチャーをセンサーの能力内で測定することになる。その結果、センサーの能力外の自然なインタラクションのふるまいを見逃してしまう危険性がある。特に、人が孤立するといった、機械が読み取ることが難しい状況は、その場に参与している人々にとっては「孤立」していることが自明のことであっても、何によって「孤立」といったラベルが会話参与者間で振られるのかが未だ不明である。

このため本研究においては、デジタルビデオカメラで撮影した立食形式パーティーの映像を使い、それにアノテーションを付与し、その情報から孤立した人を判別する方法を提案する。単純に考えて、孤立者を検出するには、単に、一定時間に発話が行われていない者を検出すればよいかもしれない。しかしながら、実際には「傍参与者(side participant)」(坊農ら[3])といった、会話を傍らで聞いている者もいるため、その方法では「傍参与者」よりさらに会話から距離のある「孤立者」をうまく検出することが難しい。

そこで本稿では、まず会話場が形成されているのかどうか、また誰が参与しているのかをアノテーションデータから判断する。こうして会話の中心部分を探し出すことにより、誰がいつ孤立しているのかがわかるようになると思われる。つまり逆転

[†] 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology
^{††} 国立情報学研究所
National Institute of Informatics

の発想で、複数のラベルから得られた会話集団の情報から、会話の中心的なアノテーションが振られない人物を「孤立者」として解釈するのである。

2. 関連研究

2.1 立食形式パーティーの分析研究

多人数インタラクション・マルチモーダルインタラクション研究に関して、日常的な会話場面のデータ収録が盛んに行われている。しかし、本研究同様の立食形式パーティーを対象とし、言語内容以外の情報を利用した分析研究は、著者が調査した範囲では以下に紹介する研究を除いて見当たらなかった。

大泉ら[4][5]は、室内などの限定された空間内における人間の行動特性を知ることが、対象となる空間のスケールやテーブルの配置などを考察する上で有用な資料となり得るという考えに基づき、限定空間内における立食パーティーを事例として、参加者の行動の時系列変化と周辺環境条件を調査した。調査は、場面 1) 研究室内の懇親交流会 (参加者数 28 名)、場面 2) クラブの OB 会 (参加者数 37 名) の 2 場面について、1) ビデオカメラによる参加者の行動記録、2) ダミーとなった参加者による観察記録、3) スチールカメラによる状況撮影記録、の 3 つのデータにより分析を行った。

その結果、参加者の行動に着目した場合、一度停留した場所に再度戻ってくる傾向が、行動軌跡から確認された。さらに、この傾向を軸にし、二度以上同じ場所に停留した位置と停留時間の関係を見ると、二度以上同じ場所に停留していた時間の総時間に対する割合は、場面 1 が約 57% (一箇所あたりの平均停留時間約 6 分)、場面 2 が約 62% (一箇所あたりの平均停留時間約 10 分) と、同じ場所に停留している時間が比較的長いことがわかり、立食パーティーの場合、各自の場が固定されず移動が自由であるにもかかわらず、参加者が特定の停留位置に何度か戻ってくるいわゆる帰巢の場が構成されることがわかった。また、参加者の停留位置の推移に着目した場合、人の立場によって停留位置の推移に違いがあることがわかった。例えば、クラブの顧問や OB 会長といった比較的ヒエラルキーの高い立場にある参加者は、来客席として用意されたテーブル周辺に停留する傾向があり、また、特定の領域を保持せずに次の停留位置に移動する傾向が見られる参加者もあり、場の設定や人の立場によって自ずから「場の優位性」が変容することが観察できた。

大泉らの研究では、本研究と同様の立食形式パーティーを対象としているが、大泉らが参加者の移動・停留について着目しているのに対し、本研究では参加者の会話場を対象とした孤立という現象に着目している点が大きく異なる。

2.2 立食パーティーにおけるコミュニケーションに関するシステム開発研究

本研究が目指す最終目標は、立食形式パーティーにおける孤立者の支援である。著者らは、本研究の成果を基礎とし、最終的には、立食パーティーにおける孤立者を判

定し、何らかの支援を行うシステムを構築することを目指している。そこで、本節では、立食パーティーに関するコミュニケーション支援システム開発研究について概観する。

大橋ら[6]は、立食パーティーのようにコミュニケーション空間内に複数の話題が並存することが可能な、「場」共有型のリアルタイムコミュニケーションシステムを提案・実装した。提案システムでは、3D サウンドを利用することでコミュニケーション空間に距離と方向の概念を導入し、現実世界と同様の複数の会話が並存できる環境を実現している。既存のデバイスや 3D サウンドを実現するライブラリの性能上、複数の会話が並存した場合会話の識別が難しくなるが、興味のある会話の音源を既存の環境で聞き取りやすい位置に移動し、視覚情報を利用して音源定位を補助するというアプローチで解決が図られている。大橋らの研究は、立食パーティーのようなコミュニケーション空間を仮想空間上でも実現可能とすることを目指した研究であり、現実世界におけるコミュニケーション空間を対象とする本研究とは、根本的に異なる。

また、Aoki ら[7]は、立食パーティーのようにコミュニケーション空間内に複数の話題が存在する場合に、複数の話題に参加するための同時的会話環境“The mad hatter's cocktail party”を提案・実装した。このシステムでは、各話者の発話区間と無音区間の相互関係に基づき、参加者の位置関係に関わらず、誰が同一の会話場に属するかを自動的に判定し、同一会話場に属する話者の発話音声により明確に聞こえるように、自動的に音響調整を行う機能を付加している。Aoki らの研究は、現実世界におけるコミュニケーション空間を対象とした点は同じであるが、本研究では、孤立した話者を支援するというより特定状況を対象としているため、本研究とは根本的に異なる。

3. データ収集実験

立食パーティーにおける孤立者を検出するためのハンドアノテーションをどのようなデザインにすればいいのか検討する上で、実際の立食パーティーを記録した動画を検証していく必要がある。そこで以下のような実験を行った。

実験会場は北陸先端科学技術大学院大学の知識科学研究科講義棟 1 階の吹き抜けになっているコミュニケーションスペースを利用した。図 1 のようにテーブルを 3 つ設置し、それぞれのテーブルを撮影するためにカメラを 3 台設置した。そして人の頭が点で見えるように撮影するために、吹き抜けの 2 階部分からポールで固定したカメラ 1 台を設置して撮影を行った。使用したデジタルビデオカメラは Victor・JVC HARD DISK MOVIE GZ-MG575 Everio である。なお吹き抜け部分から撮影を行ったカメラにはワイドコンバージョンレンズを使用し広角で撮影を行った。これにより会場全体を俯瞰で撮影することが出来た。

実験参加者数は、1 組 10 名を 2 組、合計 20 名である。設定した実験内容は「親睦

会」とし、知り合いと知り合いでない人が混在するようにした。具体的には、1組目は、3つの異なる領域の研究科からそれぞれ4, 3, 3人ずつ集め、2組目は全て同領域の研究科から、修士1年生が5人、修士2年生が5人を集めた。なお1組目と2組目の実験参加者は、重複がないようにした。「お手洗いやその他緊急の場合を除いては会場から出ない」という制限以外は、会話や行動に一切の制限は設けていない。親睦会開始直後に1人ずつ自己紹介をしてもらい、そのあとはフリートークをもらった。収録時間は2時間とした。実験は2日間に分けて行った。実験後に実験参加者に対して実験参加者同士の関係性に関するアンケートを記入してもらった。

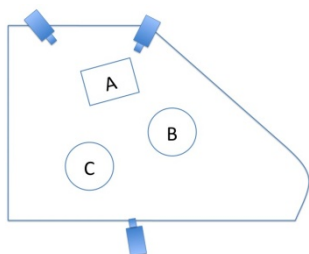


図1 実験会場

4. アノテーションデザイン

ハンドアノテーションとは、映像データに対して手作業でラベル付けを行い、データ化するアノテーション手法のことである。視線、顔の向き、発話、発話内容、身体位置、身体方向など、自分の分析に必要なと思われるレイヤー、ラベルを自由に設定することが出来る。分析対象を非常に細かくデータ化出来るのが特徴である。今回はELAN[8]というアノテーション用ソフトウェアを利用して4つの映像を同期させ、そこからハンドアノテーションを行った。レイヤーは「発話」「視線」「身体位置」「身体方向」の4つである。以下にレイヤーとそこに属するラベルについて詳細を述べる。

4.1 発話単位

今回は詳細な音声データを取得していない(各被験者にマイクを付けて録音をしておらず、ビデオカメラに内蔵されているマイクを使用して録音した)ので、正確な発話単位を取得できない。このため単純に音声が始まった点と終わった点を見だし、これらをスタートポイント、エンドポイントとした。

話者aの発話をアノテーションした例を図2に示す。横軸は時間である。まず話者aが会話を開始したのが左端である。これがスタートポイントである。そして最初の区間tの開始点がエンドポイントであり、ここまでが1発話である。区間tは、無音

区間である。発話単位ではこのように、1つの発話を細かく見ていく必要がある。発話と無音区間の時間長を計る、IPU 検出手法(伝ら[9])を用いてラベル付けを行う。しかしながら上記でもあるように今回は詳細な音声データが無いのでそこまでのラベル付けは行えないため、音声と口の動きから1発話をラベル付けしている。



図2 発話単位のラベル例

4.2 身体方向

身体方向、身体位置は共に図1のように設置されたテーブルを基準に考える。両足が地面についた状態で、かつその身体がいずれかのテーブルの方向を向いていた時にそのテーブルの記号(A-C)をラベリングする。また次の方向に身体が移動し始めた時を身体の移動の準備期間として「t」というラベルを用いる。

ラベル付けの方法を、図3の例で説明する。最初、参加者bはテーブルCの方向を向いている。その後bは、Bのテーブルに方向を変え、そのままBの方を向いたままになる。この一連の流れをアノテーションすると図4のようになる。SPとは発話者のことである。最初、bはCの方向を向いて安定しているため、「C」とラベル付けする。その後、Bに向かって方向を変えていく。この際、方向転換を開始した時点と、方向転換を終了して再度安定する時点の間が準備期間であり、やはり「t」とラベル付けする。最終的にBの方向を向いて安定し、その後Bの方向を向いている期間を「B」とラベル付けする。

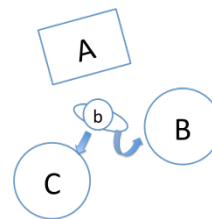


図3 身体方向の例、その1

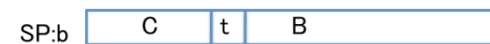


図4 身体方向のラベル例

次にどのようにラベリングしたらよいか、判断に困る場合について説明する。図5のbは単純にCの方向を向いているので「C」とラベル付けできる。しかし、aは「A」とも「C」ともとれる方向を向いている。この場合は最寄りのテーブルの記号を当てはめる。つまり図5の例では、aは現在「C」に近い、かつ両肩の方向が「C」を向いているので「C」とラベル付けする。また図6のようにbがテーブルが存在しない

方向を向いている時は「away」とラベル付けする.

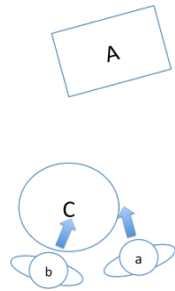


図5 身体方向の例, その2

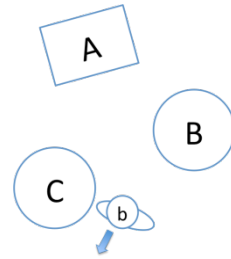


図6 身体方向の例, その3

このように, 身体方向についてはそれぞれのテーブルを基準とし, それ以外の方向を向いている時は「away」とラベル付けする. 更に「人への向き(aやc)」や「具体的なオブジェクト(テーブル以外のイス, 窓, 観賞樹など)」など細部に渡ってラベル付けを行うことは可能である. しかしながら今回の“孤立者の発見”のためにはそこまで詳細なアノテーションを行う必要はないと判断し, テーブルを基準としたラベル付けにとどめた.

4.3 身体位置

身体位置についても先程の身体方向同様にテーブルを基準としてアノテーションを行った. これも同様に両足が地面について安定している状態の時の身体位置を記述する. 準備期間のスタートポイントは, これも身体方向と同じで, 身体移動が開始され片足が動き始めた時である. エンドポイントは両足が地面についた時である.

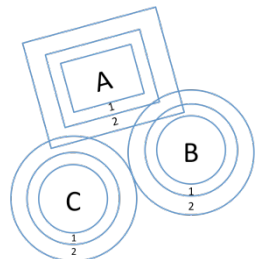


図7 身体位置のレベルの例

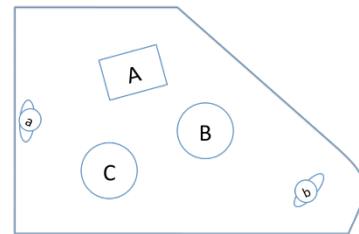


図8 身体位置の例, その1

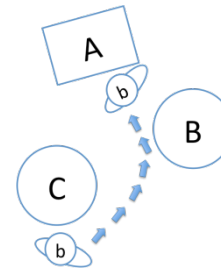


図9 身体位置の例その2



図10 身体位置のラベルの例

次に身体位置のレベルについて説明する. 図7にあるように今回はレベルを2段階とした. レベル1はもっともテーブルに接近している状態を表している. テーブルに寄りかかっておしゃべりをしたり, テーブルの上にある料理や飲み物を取るときに近づいたりした状態の時である. レベル2は, レベル1の状態の時以外の状態である. ただし, それはいずれかのテーブルに近い時である. 図8のようにどのテーブルにも近づいていない状態の場合は「none」とラベリングする. また隣接するテーブルへの移動ではなく, 図9のようにテーブルからテーブルへの距離のある移動は「Walk」とした.

実際にこの図9をアノテーションした例を図10に示す. 横方向は時間軸である. まずbがテーブルCのそばに留まっている状態「C2」からスタートし, 次に両足が地面が付いている状態から片足が少しでも離れたときから「Walk」に移行する. 「Walk」の状態は, bがテーブルAまで移動している間は継続する. bがテーブルAに接近して両足が地面について安定し, テーブルAによりかかるかあるいはテーブルA上の飲み物や食べ物を取った状態で「A1」のラベルに移行する.

図11に示す例は, 非常に良く観察されたパターンである. 特にA, B, Cを線で結んだ三角形の内側に実験参加者がいる場合にこのパターンになる場合が多い. この例で, bはテーブルA, B, Cのちょうど間にいるので「A2+B2+C2」とラベルすることが出来る. ここでbの身体方向を確認し, bがCの方向を向いているならば, 最終的なラベル付けは, Cを先頭に持ってきて「C2+A2+B2」とする.

このようにレベルごとに区分けすることと, ある程度細かい座標のラベリングをすることによって, ラベルを見るだけでその人物がどこにいるのかをおおむね判断することが出来る. 例えば「A1+B2」というラベルがあったとすると. この場合は図11でいうとbがいる位置の右斜め上のあたりにいると判断できる. より詳細なレベル分けをすればより詳細な位置がわかるようになるが, 今回はレベルは2までとした.

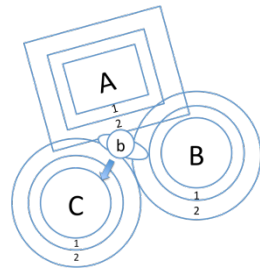


図 11 身体位置の例, その 3

4.4 視線

視線は、視線が静止している状態を基準とする。そして視線の先の「人」、「テーブル」に注目してラベリングを行う。その他、人とテーブル以外の方向に視線を向けている状態を示すための「Look up」「Look down」「Look away」をラベルとして用意した。また、視線が次の対象に向かって移動している時間を準備期間とし、やはり「t」とラベルする。

図 12 は視線の移動を一連の流れとして表した例である。ここでは a の視線についてアノテーションを行う。テーブル A の周りには、a, b, c の 3 人がいる。a ははじめに c を見ている。次に視線の移動を開始し、今度は b を見る。そして再び視線の移動を開始し、最後にテーブル A を見る。

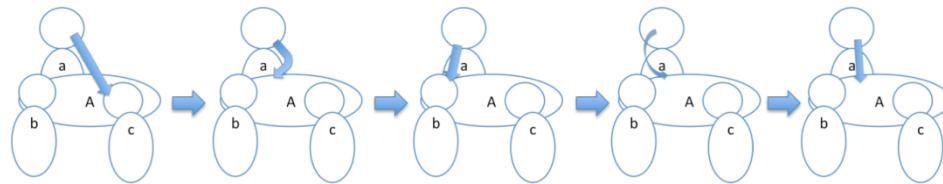


図 12 視線移動の例

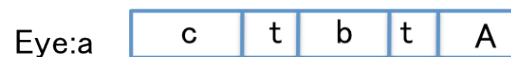


図 13 視線のラベルの例

この一連の流れをラベリングすると図 13 のようになる。横軸は時間である(実際には 1 発話中の視線移動はもっと頻繁に行われている。今回の例は見やすいようにしてあるので視線移動にかかった時間は無視して書いてある)。視線の先にある人やモノをラベルとしているので、まず「c」がくる。そして視線が移動している間は準備期間と

なるので「t」となる。あとは同じ要領で「b」「t」「A」という順にラベル付けする。

4.5 アノテーションデザインのまとめ

簡単に今までに出てきた 4 つのハンドアノテーション手法についてまとめる。まず発話単位についてだが、今回は詳細な音声データを取得していなかったため正確な発話単位を算出はしていない。単純に音声が始まった点と終わった点をスタートポイント、エンドポイントとした。発話単位以外の身体に関するアノテーションは共通した方法(準備期間(t)と具体的なラベル X)でアノテーションをした。これはジェスチャー研究における「ジェスチャー単位」[10]の考え方を利用したアノテーションで、身体が安定している状態を基準として、そして次の姿勢に移動する間を準備期間とし、姿勢が安定したところからまた別のラベルに移行するというルールでラベル付けした。身体方向、身体位置は、両足がついているのを安定した状態とし、視線は、視線が物体から離れない状態が安定した状態であるとした。

5. 孤立者検出

ハンドアノテーションから得られたデータは、発話情報(発話者、発話開始時間、終了時間)、視線情報(視線の主、視線開始時間、視線終了時間、視線先)、身体位置情報(身体位置の主、身体位置開始時間、身体位置終了時間、身体位置)、身体方向(身体方向の主、身体方向、身体方向開始時間、身体方向終了時間)の 4 つである。この 4 つの情報が実験参加者 10 人分ある。実験はそれぞれ 2 時間ずつ実施したが、そのうちから孤立者が存在すると思われる箇所を約 5 分間分切り出して分析対象データとする。このデータを使い会話集団を特定することを試みる。

5.1 会話集団

本稿における「会話集団」の定義とは、「話者」と「話者が見ている人」、「話者が見ている人を見ている人」、「話者を見ている人」および「話者との距離が近い人」という条件を満たした人々で構成される集団である(図 14 参照)。つまり参与関係にある人とさらにその外側にいる傍参与者を一つの集団と位置づけている。そして、ある時点でいずれの会話集団にも属していない人を「孤立者」とする。会話集団の形成は 1 発話ごとに行われ、その都度変化している。

5.2 会話集団に必要なデータ

孤立者を直接検出することは難しい。単純に壁ぎわにいる人や一人きりになっている人を特定することは容易であるが、しかし多人数インタラクションの場合は様々な状況が想定される。安易に孤立者を定義してしまうと、想定外の孤立者を見落としてしまう可能性がある。たとえば会話集団に紛れ込んだ孤立者と、会話集団に含まれる傍参与者を区別できなければならない。

そこで本稿では、直接孤立者を検出するのではなく、まず会話集団を特定し、いず

れの会話集団にも所属しない人物を孤立者であるとした。これにより新しい孤立者の知見が得られると考えた。発話情報だけに基づくと、発話している人とそうでない人にわけることができる。会話をしている人とそれに反応し会話を返す人、会話のキャッチボールをしている人たちを見いだすことは可能だと思われる。しかし、聞き手に徹している人を発話の情報だけでは検出することはできない。ゆえに、発話情報のみからでは会話集団を特定できない。発話情報と視線情報を組み合わせると、発話情報により誰が誰と会話しているのかがわかり、さらに視線情報からその時に誰が誰を見ているのかわかる。つまり、図 14 のように会話している人とそれを聞いている聞き手がわかる。しかしこの場合には、聞き手の情報が視線しかないの、話者から離れた位置にいる人も検出される可能性がある。たとえば図 15 のように、g が実際には会話集団に入っていない場合でも、g の視線が会話集団の誰かを見ているならば、g がその会話集団に入っていると判定されてしまう。そこでさらに発話情報、視線情報、身体位置情報を組み合わせる。発話と視線の情報から誰が発話をしていて、誰が誰を見ているのかがわかり、さらに身体位置情報を加えると、図 15 の状態でも a, b, c, d, e, f が同じテーブル A 付近にいることがわかり、g はテーブル C 付近にいることがわかる。この情報を統合し、判断すれば a, b, c, d, e, f が同一の会話集団に属していると判定できる。以上の考察により、発話情報、視線情報、身体位置情報を用いることによって会話集団の特定が可能であると考えた。

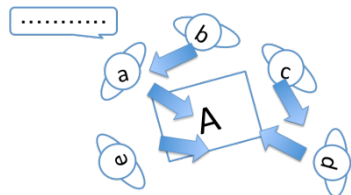


図 14 会話集団の例(発話と視線と位置)

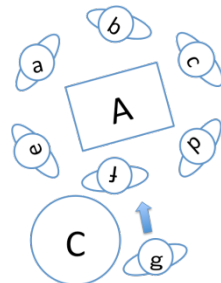


図 15 孤立者の例

5.3 孤立者検出アルゴリズム

1. 初めに、1 発話中の全視線情報を取得する。
 - 話者の発話開始時間と視線の移動開始時間、発話終了時間と視線の移動終了時間を見る。発話時間に視線の移動時間が少しでも重なるか、あるいは収まるか判定を行い 1 発話中の視線情報を取得する。
 - 取得した情報をリスト化する。これにより話者の 1 発話中に誰が、どこを見ているのか、それがどのくらいの長さなのかということがわかる。

2. 次に 1 発話中に話者が見ている人の情報を取得する。
 - 1. で得られた情報を使い、話者の視線情報を取得する。話者と視線の主を照合し、一致していかつ同一発話内であった場合はそれが話者の視線である。
 - 上記の条件にあったものをリスト化する。これにより 1 発話中の話者の視線がどこなのかわかる。
3. 1 発話中に話者が見ている人を見ている人の情報を取得する。
 - 1., 2. で得られた情報を使い、1 発話中の視線情報と話者が見ている人の情報を取得する。
 - 1 発話内の話者が見ている人と 1 発話内の全視線主の情報を照合する。この条件に合うものが話者が見ている人を見ている人である。
 - 上記の条件にあったものをリスト化する。
4. 1 発話中の話者と話者を見ている人の情報を取得する。
 - 1. で得られた情報を使い、1 発話中の視線情報と話者の情報を取得する。
 - 話者と視線情報を照合し、一致すればその時の視線の主が話者を見ている人である。
 - 上記の条件にあったものをリスト化する。
5. 1 発話中の話者、話者が見ている人、話者が見ている人を見ている人、話者を見ている人の情報を取得する
 - 2., 3., 4. のデータを結合する。
 - 上記をリスト化する。
6. 1 発話中の全身体位置情報を取得。
 - 話者の発話開始時間と身体位置開始時間、発話終了時間と身体位置終了時間を見る。
 - 発話時間に身体位置時間が少しでも被るか、あるいは収まるか判定を行い 1 発話中の身体位置情報を取得する。
 - 取得した情報をリスト化する。
7. 1 発話中の話者と 1 発話中に存在する身体位置情報、視線情報、発話時間、視線時間、身体位置時間の取得をする。
 - 1., 6. で得られた情報を使い、1 発話中の視線情報とそれに伴う身体位置情報を取得する。
 - 取得した情報をリスト化する。
8. 1 発話中の話者、話者が見ている人、話者が見ている人を見ている人、話者を見ている人それらに対応する身体位置情報、視線情報、発話時間、視線時間、身体位置時間の取得をする。
 - 5., 7. で得られた情報を使い、1 発話中の話者、話者が見ている人、話者が見ている人を見ている人、話者を見ている人それぞれに身体位置情報を添付す

る。

- 取得した情報をリスト化する。
9. 1 発話中の話者とその会話に参加している人を抽出する。
- 8.で得られたデータから話者と発話開始時間、発話終了時間を照合し、その発話内の会話参加者を取得する。
 - 話者の身体位置情報とそれぞれの身体位置情報を照合し、会話集団を特定する。
 - 取得した情報をリスト化する。

以上をまとめると、1 発話内の視線、身体位置、発話時間、視線時間、身体位置時間を利用し、話者、話者が見ている人、話者が見ている人を見ている人、話者を見ている人をそれぞれ抽出する。そしてそれぞれの身体位置情報と話者の身体位置情報を照合し、それが合致するならばそれらの人々はその1 発話内において会話集団であると判定する。

6. 評価

5 章で提案した機械的アルゴリズムとハンドアノテーションで得られたデータを使って検出した孤立者検出箇所と、データ収集実験とは別の実験参加者3人に主観で選んでもらった孤立者存在箇所の比較を行い、今回の孤立者検出アルゴリズムの評価を行った。

6.1 評価実験

3人の実験参加者が直感で孤立者がいると判断した箇所は、全部で31箇所である。このうち、全員が共通で選んだ孤立箇所が c: 1箇所、d: 1箇所、h: 2箇所、合計4箇所あった。

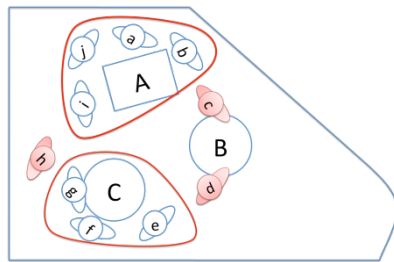


図16 孤立者存在箇所の例

図16に示した孤立者存在箇所における孤立者 c, d, h は、実験参加者3人全員が孤立していると判断した人物であり、一般的に他者から見て孤立していると判断される

状態にある。インタビューから孤立者と判断した理由をみると「視線が泳いでいた」「身体の位置が集団から離れていた」「話者との距離が離れている」「身体の向きが会話集団と違う方向を向いている」などの意見が得られた。このように、孤立者の判定には「視線」「身体位置」「身体距離」「身体方向」の4つの要素に基づいていることがわかる。その他の孤立者については、その箇所に対するインタビューから「移動時間を孤立していると考える」「違う方向に振り向く動作を孤立していると考える」などの定義付けを実験参加者が行っていたために検出された人物であることがわかった。今回はこの全ての孤立箇所を正解データとし、機械的アルゴリズムとの比較を行った。

一方、5.3 節で説明した機械的アルゴリズムを用いて1 発話毎に会話集団を見だし、どの会話集団にも属していない者を孤立者として判定する。機械的アルゴリズムを使用して検出された孤立者検出箇所は、全部で175箇所あった。

表2に、機械的アルゴリズムによって検出した孤立者検出箇所と、主観的評価によって選出された孤立者存在箇所の比較結果を示す。再現率は100%、適合率が約18%となった。つまり、機械的アルゴリズムによって得られた孤立者検出箇所は、主観的評価によって得られた孤立者存在箇所をすべて含んでいたが、主観評価では孤立者がいないと判定されていた箇所も非常に多くまれる結果となった。

	主観的孤立者	機械的アルゴリズムでの孤立者	再現率	適合率
孤立者検出箇所	31	175	100	0.177142857

表1 再現率と適合率

6.2 考察

適合率が約18%と低かった原因は3つ考えられる。第1は、1 発話中の視線の動きを用いているので、もしその発話内に視線が話者と聞き手以外の机やその他に向いている場合は会話集団に属していないとみなされ、孤立者であると判定されてしまう。第2は、現在は1 発話ごとに会話集団の判定を行っているが、無音区間については「話者がいないと」みなされ、その時点には会話集団が存在しないものと判定している。つまり、無音区間では全員が孤立者と判定されてしまう。以上の2つの理由により、無駄に多くの孤立者が検出されてしまっていると思われる。第3は、機械的アルゴリズムで検出した孤立者の中に、主観的評価では検出することができなかった「潜在的な孤立者」がいる可能性があることである。潜在的な孤立者は、図17のような状況で見受けられた。f がテーブルCにいる h と d とともに構成していた会話集団から離脱し、テーブルAの会話集団に参加しようと移動を開始した。この時、h と d はまだ会話集団を構成しているように見える。しかし実際には h と d は発話していない。さらに視線もテーブルAの会話集団やテーブルCの上を見ている。このように機械的アルゴリズムで検出した孤立者の中には、主観的に検出された孤立者に入っていない潜在

的な孤立者がいることがわかった。このような、人による主観的な評価では見いだされがたい潜在的孤立者を見いだせることは、提案アルゴリズムの重要な特長であるといえる。

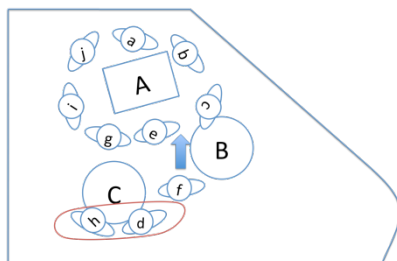


図 17 潜在的孤立者

7. おわりに

立食形式パーティーの映像データに対してハンドアノテーションを行い、そのアノテーションデータから孤立者を検出するための方法を提案した。提案手法は1発話単位で会話集団を見つけ、いずれの会話集団にも属していない人を孤立者と判定する方法である。しかし今回の提案手法では再現率は100%と高かったものの、精度が約18%と非常に低かった。その原因は1発話単位での検出では、無音区間中に全員が孤立者と判定されてしまうこと、視線が会話集団以外に向いてしまうと孤立者と判定してしまうこと、および人による主観的な評価では検出されにくい潜在的な孤立者がいたことの3点にあると考えられる。前2点は提案アルゴリズムの問題であるが、3点目の潜在的な孤立者を機械的アルゴリズムでは検出できる点は、機械的アルゴリズムの強みである。

今回提案した機械的アルゴリズムでの孤立者検出方法だけではまだ正確な孤立者の検出はできていない、今後は主観的な孤立者検出の方法と機械的アルゴリズムでの孤立者検出方法を融合させていく必要がある。特に主観的な孤立者検出と同等の方法を目指して、今回は使用していない要素である「身体距離」「身体方向」などの新しい要素を使いアルゴリズムを再構築していきたいと考えている。これにより主観的な孤立者と潜在的な孤立者の両方を検出可能になると考えている。

謝辞

本研究に際して色々なアドバイスを下さった西本研究室の皆さん、データ収集実験に参加して下さいました参加者の皆さんに感謝致します。

参考文献

- 1) Sumi, Y., Ito, S., Matsuguchi, T., Fels, S., Iwasawa, S., Mase, K., Kogure, K., and Hagita, N. Collaborative capturing, interpreting, and sharing of experiences, *Personal and Ubiquitous Computing*, Vol. 11, No. 4, pp. 213-328, (2007).
- 2) 角 康之, 伊藤 禎宣, 松口 哲也, Sidney Fels, 間瀬 健二: 協調的なインタラクションの記録と解釈, 情報処理学会論文誌, Vol. 44, No. 11, pp. 2628-2637, (2003).
- 3) 坊農 真弓, 高橋 克也: 多人数インタラクションの分析手法 (知の科学), オーム社, ISBN : 978-4-274-20732-7, (2009).
- 4) 大泉信明, 若井正一, 松下信禎: 立食パーティー会場における参加者の行動特性: 場の変容からみた人間の行動特性に関する事例研究, 学術講演梗概集. E, 建築計画, 農村計画 1994, pp. 1081-1082, 1994.
- 5) 大泉信明, 若井正一: 場の変容からみた人間の行動特性に関する事例研究: 立食パーティー会場における参加者の行動特性(建築計画系・都市計画系), 日本建築学会東北支部研究報告集, vol. 57, pp. 133-136, 1994.
- 6) 大橋純, 広淵崇宏, 河合栄治, 藤川和利, 砂原秀樹: 視覚情報により強化された3Dサウンド場による共有型多人数音声チャットシステムの設計と実装, 情報処理学会研究報告. CSEC, [コンピュータセキュリティ] 2006(26), pp. 227-232, 2006.
- 7) Paul M. Aoki, Matthew Romaine, Margaret H. Szymanski, James D. Thornton, Daniel Wilson, Allison Woodruff: The mad hatter's cocktail party: a social mobile audio space supporting multiple simultaneous conversations. CHI 2003, pp. 425-432, 2003.
- 8) ELAN: <http://www.lat-mpi.eu/tools/elan>
- 9) 伝康晴・小磯花絵・丸山岳彦・前川喜久雄・高梨克也・榎本美香・吉田奈央: 対話研究にふさわしい発話単位の認定に向けて. 人工知能学会研究会資料, SIG-SLUD-A802, pp. 27-32
- 10) 細馬宏通: 非言語コミュニケーションのための分析単位 -ジェスチャー単位-. 人工知能学会誌, 23, 390-396. (2008).